Introduction to Bayesian inference

End of course data analysis project

—

Vaccination against Varicella

The Center for Disease Control (CDC) reports the vaccination coverage of Varicella among young children. Varicella, commonly known as chicken- pox, is a highly contagious viral infection caused by the varicella-zoster virus (VZV). Vaccination against chickenpox has been highly effective in reducing the incidence and severity of the disease. In the United States, vaccination against varicella has been part of the routine childhood immunization sched- ule since the mid-1990s. Since the vaccine's introduction, there has been a dramatic decline in the number of chickenpox cases, hospitalizations, and deaths associated with the disease. The target for vaccination coverage of varicella (chickenpox) in the United States, as set by the Centers for Dis- ease Control and Prevention (CDC), is typically around 90% or higher for children. This high coverage rate is aimed at achieving herd immunity and preventing outbreaks of chickenpox within communities.

Project 1: Insurance

The next table summarizes, based on a survey, the number of children in the birth cohort 2014-2017 that had at least one dose of the Varicella vaccine. It gives the number of vaccinated children (Vaccinated) amongst the number of children in the survey (Sample Size). The information is provided for 5 regions of the US, and split according to insurance status (private insurance, uninsured or any Medicaid).

Geography	Insurance	Vaccinated	Sample Size
North Carolina	Any Medicaid	380	419
North Carolina	Private Insurance Only	632	673
North Carolina	Uninsured	28	34
Georgia	Any Medicaid	363	396
Georgia	Private Insurance Only	527	576
Georgia	Uninsured	36	50

Geography	Insurance	Vaccinated	Sample Size
Wisconsin	Any Medicaid	282	332
Wisconsin	Private Insurance Only	514	548
Wisconsin	Uninsured	16	34
Florida	Any Medicaid	446	490
Florida	Private Insurance Only	588	628
Florida	Uninsured	28	39
Mississippi	Private Insurance Only	400	441
Mississippi	Uninsured	27	32

Question 1

> Derive analytically the posterior of the vaccination coverage per ge- ography and insurance group. Use a conjugate prior that (1) reflects no knowledge on the vaccination coverage, and (2) reflects that vac- cination coverage is typically around 90% or higher. Give posterior summary measures of the vaccination coverage per geography and in- surance group. Is the choice of the prior impacting your results?

Theoretical considerations

The outcome *Vaccinated/Not Vaccinated* follows a Bernouilli distribution with parameter p:

V : Vaccination status $V \in \{0, 1\}$ $V \sim \mathcal{B}ern(p)$

It is known from theory that the sum of n *i.i.d* Bernoulli random variables follows a Binomial distribution. This will be used to model the sample outcome: the number of vaccinated people V_s in a random sample of size n:

$$V_s = \sum_{i}^{n} V_i \sim \mathcal{B}inom(n, \theta)$$

where θ is the parameter of interest - the vaccine coverage.

In the course, we saw that the Beta distribution is the conjugate prior for binomially distributed data:

Distribution	Formula
Prior	$p(\theta) = \mathcal{B}eta(\alpha, \beta)$
Likelihood	$p(y \mid \theta) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$
Posterior	$p(\theta \mid y) = \mathcal{B}eta(\alpha+k, \beta+n-k)$

The summary measures for the Beta distribution are defined as follows:

Summary Measure	Formula
Mean	$\frac{\alpha}{\alpha+\beta}$
Median	See Note
Mode	$\frac{\alpha-1}{\alpha+\beta-2}$ for $\alpha, \beta > 1$

Note: The median of the Beta distribution does not have a simple closed form expression. It can be approximated numerically or using statistical software.

```r
source("./project.R")
```

```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union


Attaching package: 'kableExtra'

The following object is masked from 'package:dplyr':

    group_rows
```

Choice of prior distributions

(1) No prior knowledge

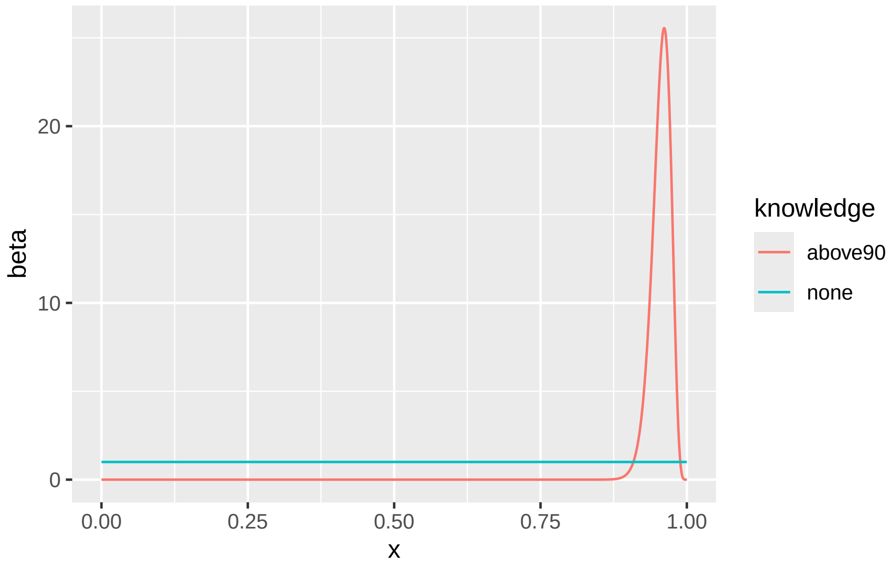
In order to reflect no prior knowledge on the vaccine coverage, the weakly-informative prior Beta(1,1) will be used, which is equivalent to the uniform distribution over $[1, 1]$.

(2) Vaccine coverage >90%

For modeling prior knowledge that vaccine coverage is about 90%, we chose the Beta(150, 7) distribution.

Comparison of priors

```r
plot_priors
```



Results

```
uni_posterior %>% knitr::kable(
    caption = "Posterior distribution parameters and summary measures per geography and insu
    col.names = c(
        "Geography", "Insurance",
        "alpha", "beta", "mean", "mode"
    ),
    digits = 2) -> testtbl
testtbl %>%
  add_header_above(header = c(
    " "=2, "Beta(1,1) prior"=4
  ))
```

Table 4: Posterior distribution parameters and summary measures per geography and insurance for Beta(1,1) prior

		Beta(1,1) prior			
Geography	Insurance	alpha	beta	mean	mode
North Carolina	Any Medicaid	381	40	0.90	0.91
North Carolina	Private Insurance Only	633	42	0.94	0.94
North Carolina	Uninsured	29	7	0.81	0.82
Georgia	Any Medicaid	364	34	0.91	0.92
Georgia	Private Insurance Only	528	50	0.91	0.91

Georgia	Uninsured	37	15	0.71	0.72
Wisconsin	Any Medicaid	283	51	0.85	0.85
Wisconsin	Private Insurance Only	515	35	0.94	0.94
Wisconsin	Uninsured	17	19	0.47	0.47
Florida	Any Medicaid	447	45	0.91	0.91
Florida	Private Insurance Only	589	41	0.93	0.94
Florida	Uninsured	29	12	0.71	0.72
Mississippi	Private Insurance Only	401	42	0.91	0.91
Mississippi	Uninsured	28	6	0.82	0.84